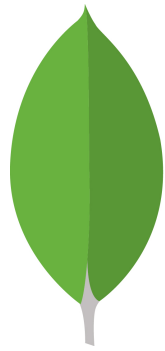


Replicación



mongoDB

Concepto

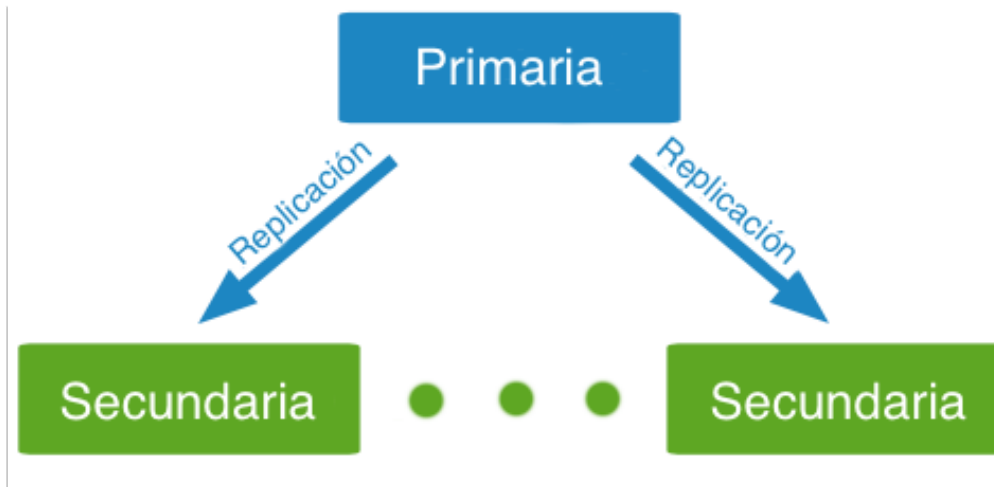
- Objetivo: incrementar la redundancia de datos para alta disponibilidad e integridad.
- Una base de datos al tener varias copias exactas en diferentes infraestructuras separadas asegura que si una de las instancias falla, ya sea una falla catastrófica a nivel de hardware o situaciones diversas que pudiesen corromper y/o evitar el acceso a la data, el sistema que dicha base de datos alimente no se vea afectado ya que existen otras instancias **espejo** que tomarán el lugar de la original mediante un proceso **transparente para los usuarios** finales.

Arquitectura

- Al grupo de instancias que poseerán la misma información se les denomina **replica set** o grupo de replicación.
- Un **replica set** en MongoDB está compuesto por 2 tipos principales de miembros, instancias **primarias** y **secundarias**:
 - Una única instancia **primaria** la cual aceptará todas las operaciones de escritura provenientes de los sistemas cliente.
 - Es posible configurar la lectura a instancias secundarias.

Arquitectura (2)

- Las operaciones que alteran los datos son escritas en un archivo llamado **oplog** o bitácora de operaciones, los miembros secundarios replican este archivo del miembro primario y ejecutan las mismas operaciones sobre su conjunto de datos, esto permite tener la misma información en las diferentes instancias.



Debemos tomar en cuenta que debido a que el procedimiento de replicación se realiza de manera asíncrona, es posible que clientes que consulten directamente a miembros secundarios no obtengan la información más reciente.

Tipos de miembros secundarios

- **Miembro de prioridad 0 (NO PRIMARIO)**

Se le denomina a un miembro secundario configurado con prioridad 0, esto evitará que dicho miembro pueda ser elegido a convertirse en miembro primario en caso de que aquel falle.

- **Miembro escondido (NO PRIMARIO NI ACCESO)**

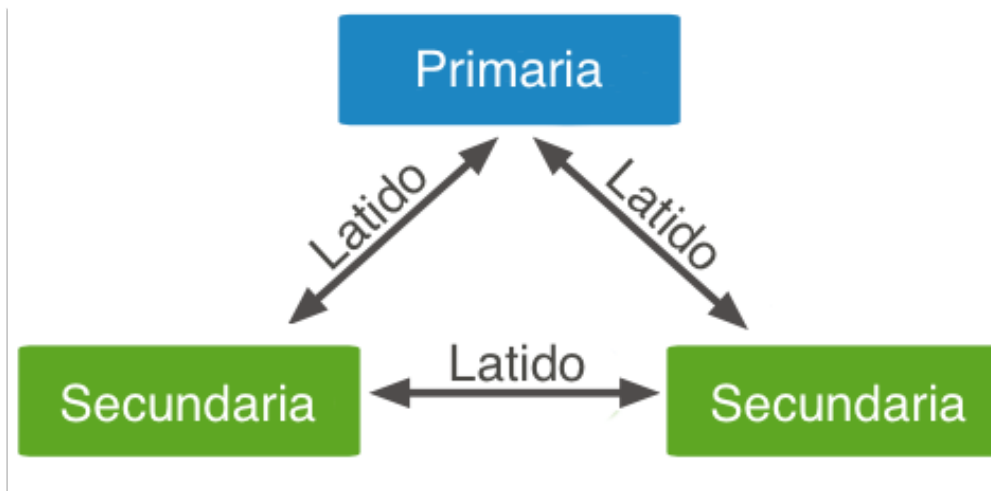
Son miembros secundarios de tipo prioridad 0 pero que además se les niega la posibilidad de ser accesibles para lectura por parte de los sistemas cliente.

- **Miembro retrasado**

También son miembros de prioridad 0 y poseen la cualidad particular de mantener un estado retrasado de la base de datos, suele utilizarse como instancias de respaldo ya que no han sido afectadas por las últimas operaciones que pudiesen estar alterando de manera no deseada la información. Debido al estado retrasado de este miembro se recomienda que también se defina como un miembro escondido.

Elección

- La arquitectura de los **replica set** dicta que los miembros deben enviar latidos o *pings* entre ellos cada 2 segundos, si en un período de 10 segundos el latido no es devuelto, se marca al miembro en cuestión como inaccesible.



Llamar a elección

- Una elección se lleva a cabo cuando no existe un miembro primario, este deja de responder o este es obligado a darse de baja, veamos las diferentes condiciones:
 - La iniciación de un nuevo **replica set**.
 - Un miembro secundario pierde contacto con el primario.
 - El miembro primario es obligado a convertirse en secundario.
 - Si un secundario es elegible para elecciones y posee un mayor índice de prioridad.

Prioridad

- La elección de un nuevo miembro primario se basa inicial y principalmente en la comparación de prioridades de aquellos miembros elegibles, esta prioridad es por defecto 1, esto para darles a todos los miembros la posibilidad de ser elegidos.
- Si un miembro de mayor prioridad alcanza a tener al día su información con al menos 10 segundos con respecto al **oplog**, se declara una elección para darle a dicho nodo de mayor prioridad la oportunidad de convertirse en primario.

Optime

- Se toma en cuenta la estampilla de tiempo más reciente de la última operación que el miembro aplicó del **oplog**, por ello se llama **optime** o tiempo de operación.

Conectividad

- Otro aspecto a considerar es la capacidad que tiene el candidato para conectarse con la mayoría de de los miembros en el grupo.
- En este caso particular se le llama "miembros" a aquellos que poseen la capacidad para votar. Al configurar un miembro del **replica set** se puede indicar si a este se le permite votar o no y cuantos votos posee cada uno.

Negaciones en elecciones

- Cualquier miembro del **replica set** puede vetar o negar una elección (incluyendo aquellos miembros que no votan) si ocurre alguna las siguientes situaciones:
 - Si el miembro que busca la elección no es parte del grupo de votantes.
 - Si el miembro que busca la elección no está al día con las operaciones más recientes en el **oplog** que son accesibles por el **replica set**.
 - Si el miembro que busca la elección posee una prioridad menor a algún otro miembro que también sea elegible.

Miembros impares

- Procura que la cantidad de miembros en tu **replica set** sea impar, esto facilitará los procesos de elecciones y solicitará menos recursos y menos tiempo.
- Para estos casos es posible configurar un miembro **árbitro**. Este tipo de miembros **no** poseen una copia del conjunto de datos y por ende no pueden volverse primarios. Pueden ser de gran utilidad cuando **no posees la infraestructura** para soportar un miembro más pero deseas seguir el estándar de miembros impares ya que la cantidad de recursos que necesita es mucho menor.

Número de miembros

- Un **replica set** puede tener un máximo de 12 miembros y solo 7 de ellos con la capacidad de votar.
- En caso de que tu solución necesite un mayor número de miembros es posible implementar la estrategia precedente a los **replica set** conocida como replicación maestro-esclavo.
- Se debe tomar en cuenta que esta estrategia ofrece menos redundancia y no soporta el proceso automático de recuperación a fallas como lo hacen los **replica set** con los latidos y elecciones.

Distribución geográfica



Distribución geográfica

- Procure subdividir los miembros de la **replica set** en distintos centros de datos, esto asegurará que si algo le sucede al centro de datos, la base de datos se mantendrá activa.
- Trate de mantener una mayoría en uno de los centros para garantizar que los miembros puedan elegir a un primario en caso de que la comunicación entre otros miembros localizados en otro centro se vea dificultado.
- Inclusive se puede tener un miembro tipo prioridad 0 en un centro de datos aparte como respaldo.